Virtual Screening Workflow

Schrödinger Suite 2006

Schrödinger Press

Copyright © 2006 Schrödinger, LLC. All rights reserved. CombiGlide, Epik, Glide, Impact, Jaguar, Liaison, LigPrep, Maestro, Phase, Prime, QikProp, QikFit, QikSim, QSite, SiteMap, and Strike are trademarks of Schrödinger, LLC.
Schrödinger and MacroModel are registered trademarks of Schrödinger, LLC.

Python is a copyrighted work of the Python Software Foundation. All rights reserved.

The C and C++ libraries for parsing PDB records are a copyrighted work (1989) of the Regents of the University of California. All rights reserved.

See the Copyright Notices for full copyright details.

To the maximum extent permitted by applicable law, this publication is provided "as is" without warranty of any kind. This publication may contain trademarks of other companies.

Please note that any third party programs ("Third Party Programs") or third party Web sites ("Linked Sites") referred to in this document may be subject to third party license agreements and fees. Schrödinger, LLC and its affiliates have no responsibility or liability, directly or indirectly, for the Third Party Programs or for the Linked Sites or for any damage or loss alleged to be caused by or in connection with use of or reliance thereon. Any warranties that we make regarding our own products and services do not apply to the Third Party Programs or Linked Sites, or to the interaction between, or interoperability of, our products and services and the Third Party Programs. Referrals and links to Third Party Programs and Linked Sites do not constitute an endorsement of such Third Party Programs or Linked Sites.

May 2006

Document Conventions

In addition to the use of italics for names of documents, the font conventions that are used in this document are summarized in the table below.

Table 1.1.

Font	Example	Use
Sans serif	Project Table	Names of GUI features, such as panels, menus, menu items, buttons, and labels
Monospace	`$SCHRODINGER/maestro`	File names, directory names, commands, environment variables, and screen output
Italic	*filename*	Text that the user must replace with a value
Sans serif uppercase	CTRL+H	Keyboard keys

In descriptions of command syntax, the following UNIX conventions are used: braces `{ }` enclose a choice of required items, square brackets `[ ]` enclose optional items, and the bar symbol | separates items in a list from which one item must be chosen. Lines of command syntax that wrap should be interpreted as a single command.

In this document, to *type* text means to type the required text in the specified location, and to *enter* text means to type the required text, then press the ENTER key.

References to literature sources are given in square brackets, like this: [10].

Virtual Screening Workflow

The Virtual Screening Workflow is designed to run an entire sequence of jobs for screening large collections of compounds against a particular target. The workflow includes ligand preparation using LigPrep, filtering using `propfilter` on QikProp properties or other structural properties, and Glide docking at the three accuracy levels, HTVS, SP, and XP. The design allows you to choose which of the stages to include in any run, and which selection of results from one stage are passed on to the next.

Before you run the workflow, you must ensure that receptor is properly prepared, and you must generate Glide grids for the receptor. See Chapter 4 of the *Glide User Manual* for information on protein preparation, and Chapter 6 of the *Glide User Manual* for information on grid generation.

The ligand files for the workflow can consist of 2D structures or 3D structures. If you supply 2D structures, you must run the ligand preparation part of the workflow to convert the structures to 3D for docking. This stage runs a LigPrep job, with the standard options for most parts of the LigPrep process. You can control some of these options, but if you want to use other options, you should run LigPrep on the ligands independently. For more information, see the *LigPrep User Manual*.

Prefiltering of the ligands can also be performed as part of the workflow. If you want to use the Lipinski Rule option, you must run QikProp to obtain the required properties. You can run QikProp as part of the workflow, regardless of the ligand source. If you have already run QikProp, you do not need to run it again. If you want to specify a custom filter, you can provide an input file for `propfilter`, which will then filter the structures. For more information on `propfilter`, see Section D.2 of the *Maestro User Manual*.

The full workflow includes three docking stages. The first stage performs HTVS docking. The ligands that are retained are then passed to the next stage, which performs SP docking; the survivors of this stage are passed on to the third stage, which performs XP docking. At each stage you can decide how many ligands are kept, and whether all the ionization and tautomeric states of each ligand are kept, or only some.

The workflow is intended to be robust, so that if a subjob fails, the master job will attempt to rerun the job a few times before quitting. If the master job fails (for example, due to a system failure or network failure) you can restart the job, and it will pick up the calculation from the latest point for which it has results and can restart from them.

The results of multiple runs of the workflow can be merged using the script `$SCHRODINGER/utilities/glide_ensemble_merge`. Use the `-h` option for the syntax of this script.

The Virtual Screening Workflow Panel

The Virtual Screening Workflow panel sets up the input files for LigPrep, QikProp, and Glide ligand docking and submits them to the selected host in order. It also provides capabilities for monitoring and restarting the jobs if they fail.

To open the Virtual Screening Workflow panel, choose Virtual Screening Workflow from the Applications menu in the main window.

The main part of the Virtual Screening Workflow panel consists of four tabs for setting up and monitoring virtual screening jobs. The features of these tabs and how to use them are described in separate sections.

- General Tab
- Docking Options Tab
- Job Options Tab
- Monitor Jobs Tab

Below the tabs is a row of buttons. The Start button starts the job; the Write button writes the input file but does not start the job.

General Workflow Setup

In the General tab you activate the parts of the workflow that you want to perform—ligand preparation, running QikProp, filtering, and docking—specify the input sources, and set options for ligand preparation.

In the top section of the General tab, you specify the file that contains the ligands to be docked, and ensure that each ligand is uniquely identified. In the Prepare ligands section, you specify options for LigPrep. The Filtering section allows you to run QikProp and filter the ligands by property. Finally, the Perform Glide docking section allows you to specify the grid for docking.

Specifying the Input Ligand Files

To specify a single ligand file, you can enter a file name in the Input ligand file text box, or click Browse to navigate to the file.

If you want to specify multiple ligand files with related names, you can use the wild card characters `*` and `?` in the file name. These characters have their usual Unix file-matching meanings: `?` matches a single character, and `*` matches zero or more characters.

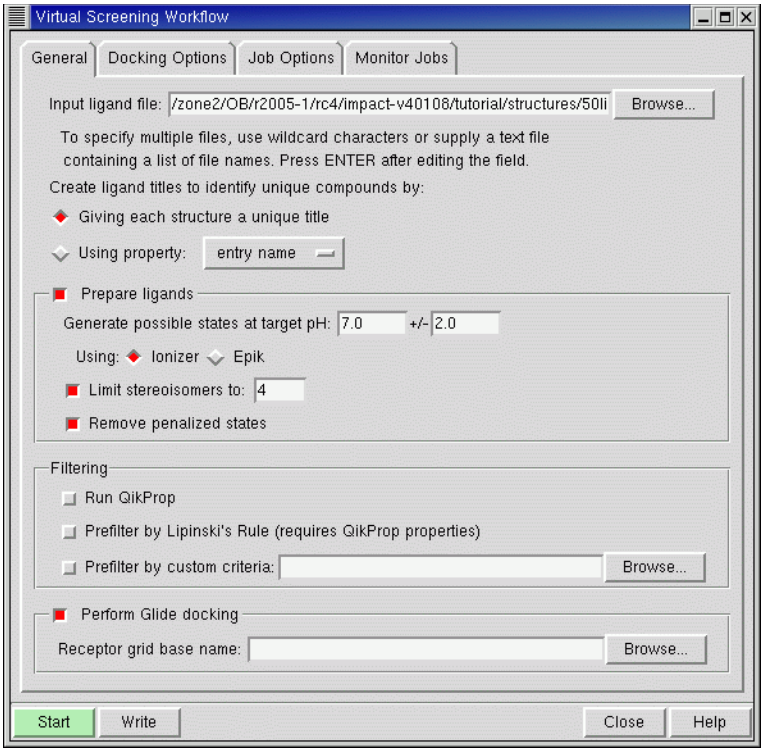

Figure 1. The** General **tab of the** Virtual Screening Workflow **panel.

To specify multiple ligand files with any kind of name, you can create a text file that contains a list of ligand file names, and specify this text file in the Input ligand file text box, or click Browse and navigate to it.

If you type in the file name, you must press ENTER to ensure that the name is read and the Using property option menu is populated.

Identifying Ligands and their States

In the docking sequence, you can choose to keep all the ionization and tautomeric states of a given compound for which one of these states docks well. The ionization and tautomeric states that originate from the same compound are identified by their title. It is therefore necessary to set the title property. The controls under Create ligand titles to identify unique compounds by enable you to set or select the title for the ligands. The controls for setting the title are not available until you have specified the ligand files.

If the structures you have are all unique compounds, you can assign a unique title to each with the first option, Giving each structure a unique title. The title that is assigned is an integer.

If the structures contain different ionization states or tautomers of the same compound, you can assign a title by selecting Using property and choosing a property from the option menu. The property names are taken from the first structure in each file, and only those properties that exist in each file are presented. You should ensure that the property you choose exists for each structure in the file, not just the first. The option menu becomes available when a file with valid properties is specified.

When the title is set, a new property is created to store the original title.

Preparing Ligands with LigPrep

If your structures are already 3D all-atom structures with the appropriate ionization and tautomeric states, you do not need to prepare the ligands, and you can deselect the Prepare ligands option. Otherwise, ensure that this option is selected, and choose the desired options for generation of states for each ligand.

To generate ionization (protonation) states that are likely to exist in a given pH range, enter the target pH and the range in the Generate possible states at target pH text boxes. You can generate these states with either the Ionizer or Epik, by selecting the appropriate option. Epik is licensed separately from LigPrep; the Ionizer is included with LigPrep. In addition to varying the ionization state, LigPrep generates tautomeric states that are likely in the given pH range. You can remove ionization and tautomeric states that have large penalties by selecting Remove penalized states. These are states that are likely to have low populations at the prevailing conditions.

LigPrep will also vary the stereochemistry, if it is not predetermined. You can limit the number of stereoisomers by entering the maximum number in the Limit stereoisomers to text box. The LigPrep run preserves any existing chirality information in the input file.

Filtering Ligands Prior to Docking

The Virtual Screening Workflow offers two choices for prefiltering ligands based on computed properties: Lipinski's Rule of 5, and a custom filter. Ligands that do not meet the specified criteria are removed from the ligand list for docking. Both prefiltering choices use propfilter, which can filter the structures in a Maestro file based on any property in the file.

To prefilter the ligands using Lipinski's Rule of 5 before docking, select Prefilter by Lipinski's Rule. Ligands that do not satisfy this rule are not docked. This option requires QikProp properties. If the input structure files do not have QikProp properties, select Run QikProp in this tab.

To choose your own filter, select Prefilter by custom criteria. You must supply the name of the input file for `propfilter`, either by entering it into the text box or by clicking Browse and navigating to the file. For more information on propfilter, see Section D.2 of the *Maestro User Manual*.

Setup for Glide Docking

If you want to dock the ligands that you have prepared and filtered, ensure that Perform Glide docking is selected, and specify the grid for the receptor. You can enter the path to the grid file (`.grd`) in the Receptor grid base name text box, or click Browse and navigate to the grid file. The name you enter should be the base name, without the `.grd` extension.

Setting Docking Options

If you selected docking as part of the workflow, you can choose which of the three docking accuracy levels to include and set various options for docking in the Docking Options tab. The docking options are described in detail in Chapter 7 of the *Glide User Manual*.

Setting Ligand Parameters

At the top of the Docking Options tab are two text boxes that specify cutoffs for the size of the ligands and the number of rotatable bonds. Ligands larger than the specified number of atoms or with more rotatable bonds than the number specified will not be docked. See Section 7.3.2 of the *Glide User Manual* for more information.

You can also soften the nonpolar part of the ligand potential, by scaling the the van der Waals radii of ligand atoms with small partial charges. To do so, enter the scaling factor and the partial charge cutoff in the text boxes in the Scaling of ligand van der Waals radii section. See Section 7.3.3 of the *Glide User Manual* for more information.

Setting Up the Docking Steps

The lower part of the Docking Options tab allows you to choose which Glide docking runs to include in the workflow, and contains options for HTVS, SP, and XP docking runs. To include a docking run in the workflow, select the corresponding option. Under each option is a set of controls for the docking stage. The same controls are provided for each option, and are described below.

The two Keep options allow you to specify the percentage (upper text box) or number (lower text box) of the best compounds to keep. A compound may consist of several ionization or tautomeric states. The option menu allows you to choose how many ionization or tautomeric states to keep for each compound.

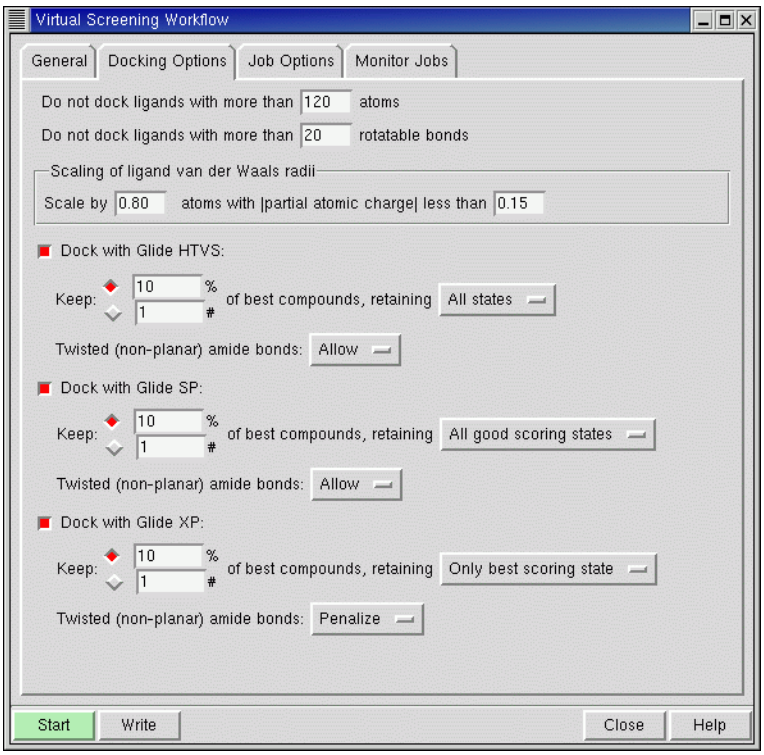

Figure 2. The Docking Options ***tab of the*** Virtual Screening Workflow ***panel.***

The options to choose from are:

- All states (default for HTVS)
- All good scoring states (default for SP)
- Only best scoring states (default for XP)

The rationale for keeping all ionization and tautomeric states is that the actual state that scores best can vary with the accuracy level. Keeping all states of a particular compound in the early stages ensures that the structures that will score best in later stages are not discarded.

The Twisted (non-planar) amide bonds option menu enables you to specify whether non-planar amide bonds are allowed in HTVS and SP docking, or penalized in XP docking.

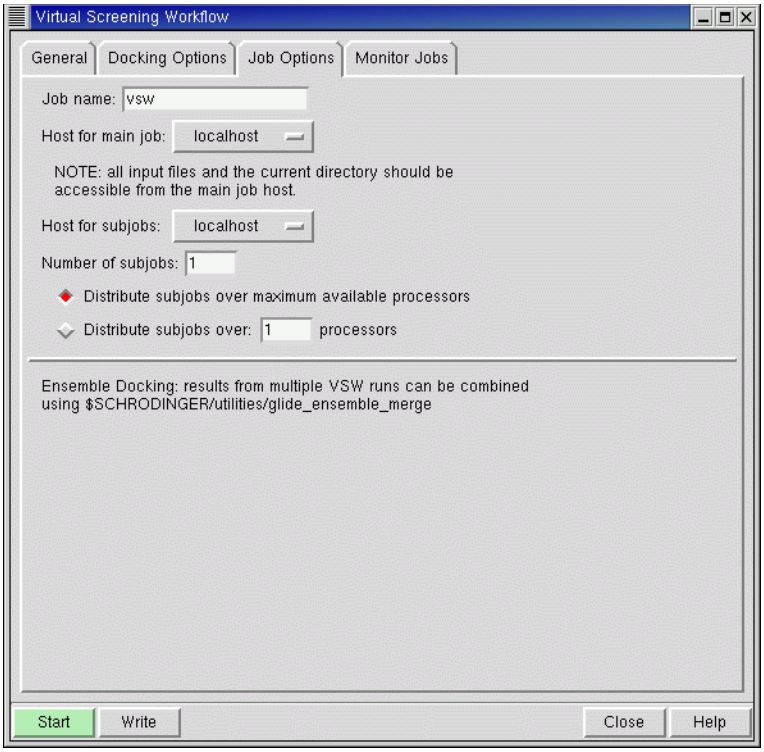

Figure 3. The Job Options ***tab of the*** Virtual Screening Workflow ***panel.***

Setting Job Options

When the general setup and the docking setup has been completed, you can set job options in the Job Options tab and start the job by clicking Start. The run consists of a main job and a set of subjobs. The main job starts all the subjobs for the various stages of the workflow, and collects the results.

You can choose separate hosts for the main job and the subjobs. The host you choose for the main job must have access to the input files and the current directory, but the host you choose for the subjobs does not need access to these files and directory.

Each stage of the workflow can be divided into a set of subjobs, which can be run concurrently. You can specify the number of subjobs in the Number of subjobs text box. You should consider all stages of the workflow when deciding on the number of subjobs. The main job may adjust the number of subjobs to better balance the load for each part of the workflow, so the number you enter is a target, not a requirement.

If you run multiple subjobs you can select a multiprocessor host such as a cluster from the Host for subjobs option menu. This host need not be the same as the main job host. You can also choose to limit the number of processors allocated to the subjobs by entering the maximum number in the Distribute subjobs over *n* processors text box. The number of processors should not be more than the number of subjobs. Otherwise, you can ensure that the maximum number of processors available is allocated to the execution of the subjobs by selecting Distribute subjobs over maximum available processors.

Monitoring and Managing Jobs

When the job is running, you can monitor and control jobs started from the Virtual Screening Workflow panel in the Monitor Jobs tab. As well as displaying information from the master job log file, you can pause, resume, kill, restart, and rerun jobs, using the buttons below the text pane. The function of these buttons is described in Table 1.

By default, the current job is selected, and is indicated by the presence of the input file name in the Job input file text box. To monitor another VSW job, enter the name of the input file for the job in this text box, or click Browse and navigate to the file. The name of the input file should be *jobname*`.inp`. When you have selected the input file, the master job ID and the job status is displayed in the panel, immediately below this text box.

You can display either the log file for the master job (main job) or the input file for the job in the text pane in the center of the tab by selecting the appropriate View option.

Jobs can also be monitored and controlled from the Maestro Monitor panel, but the Monitor Jobs tab provides in addition the capability for restarting a job. The subjobs in the restarted job resumes at the last point for which information is available for restarting.

Table 1. Function of buttons in the Monitor Jobs tab.

Button	Action
Update	Update the job status and redisplay the latest version of the log file.
Pause	Pause the selected job.
Resume	Resume execution of the selected job.
Kill	Kill the selected job.
Restart	Restart the selected job. This button is unavailable if the job is running. When a job is restarted, the subjobs of this job resume from the most recent point possible; however, it is likely that some work will be lost.
Rerun	Rerun the selected job from the beginning. This button is unavailable if the job is running. This action overwrites the files from the previous run.

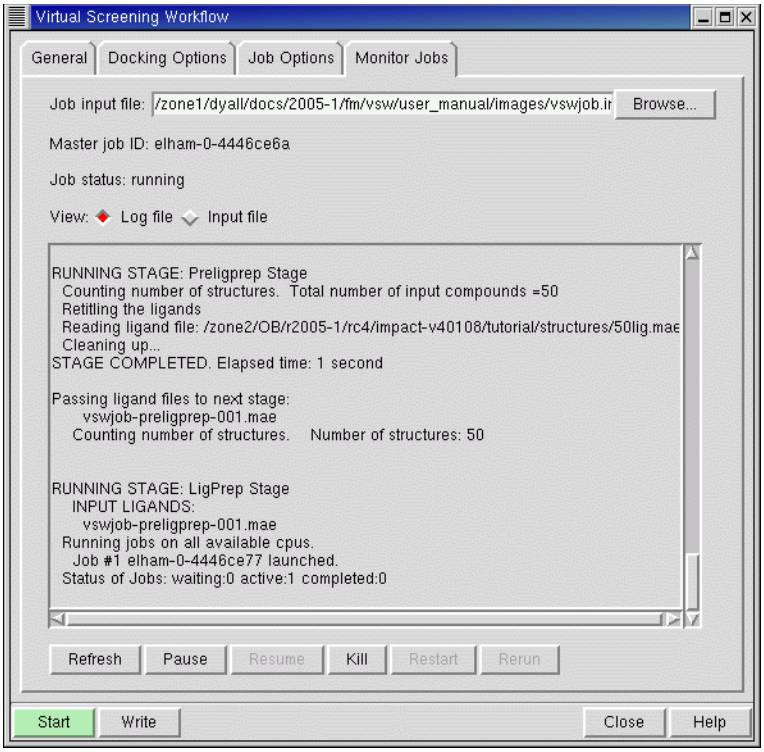

Figure 4. The Monitor Jobs ***tab of the*** Virtual Screening Workflow ***panel.***

Merging Results of Multiple Jobs

You can merge the results of several runs of the workflow into a single Glide pose viewer file, using the utility `glide_ensemble_merge`. In the process, you can specify a scoring offset for calibration of different runs. The results need not be from a single receptor: you can merge results for multiple receptors. The script combines sorted Glide pose viewer (`*_pv.mae`) files into an output file, sorted by GlideScore.

The syntax is as follows:

```
$SCHRODINGER/utilities/glide_ensemble_merge [options] {-j jobname|
    --jobname=jobname} job1_pv.mae[:offset1] job2_pv.mae[:offset2]
    [job3_pv.mae[:offset3] ... ]
```

The options are listed in Table 2. The list of files to be merged is a blank-separated list of pose-viewer file names. Each file name can be followed by a colon and an offset. The default offset is zero.

Table 2. Options for the glide_ensemble_merge utility.

Option	Description
`--version`	Display program version number and exit
`-h`, `--help`	Display usage message and exit
`-n` *maxlig*, `--nreport=` *maxlig*	Maximum number of best ligands to keep after merging. Default: 10000.
`-j` *jobname*, `--jobname=` *jobname*	Job name. Required. The output file is named *jobname*`_pv.mae`.

Citing the Virtual Screening Workflow in Publications

Schrödinger Suite 2006 Virtual Screening Workflow; Glide version 4.0, Schrödinger, LLC, New York, NY, 2006; LigPrep version 2.0, Schrödinger, LLC, New York, NY, 2006; QikProp version 2.5, Schrödinger, LLC, New York, NY, 2006.

Getting Help

Schrödinger software is distributed with documentation in PDF format. If the documentation is not installed in `$SCHRODINGER/docs` on a computer that you have access to, you should install it or ask your system administrator to install it.

For help installing and setting up licenses for Schrödinger software and installing documentation, see the *Installation Guide*. For information on running jobs, see the *Job Control Guide*.

Maestro has automatic, context-sensitive help (Auto-Help and Balloon Help, or tooltips), and an online help system. To get help, follow the steps below.

- Check the Auto-Help text box, which is located at the foot of the main window. If help is available for the task you are performing, it is automatically displayed there. Auto-Help contains a single line of information. For more detailed information, use the online help.
- If you want information about a GUI element, such as a button or option, there may be Balloon Help for the item. Pause the cursor over the element. If the Balloon Help does not appear, check that Show Balloon Help is selected in the Help menu of the main window. If there is Balloon Help for the element, it appears within a few seconds.
- For information about a panel or the folder that is displayed in a panel, click the Help button in the panel. The Help panel is opened and a relevant help topic is displayed.
- For other information in the online help, open the Help panel and locate the topic by searching or by category. You can open the Help panel by choosing Help from the Help menu on the main menu bar or by pressing CTRL+H.

If you do not find the information you need in the Maestro help system, check the following sources:

- *Glide User Manual*, for information on Glide
- *LigPrep User Manual*, for information on LigPrep
- *QikProp User Manual*, for information on QikProp
- *Maestro User Manual*, for detailed information on using Maestro
- Frequently Asked Questions pages on the Schrödinger Support Center

The manuals are also available in PDF format from the the Schrödinger Support Center. Information on additions and corrections to the manuals is available from this web page.

If you have questions that are not answered from any of the above sources, contact Schrödinger using the information below.

E-mail: help@schrodinger.com
USPS: 101 SW Main Street, Suite 1300, Portland, OR 97204
Phone: (503) 299-1150
Fax: (503) 299-4532
WWW: http://www.schrodinger.com
FTP: ftp://ftp.schrodinger.com

Generally, e-mail correspondence is best because you can send machine output, if necessary. When sending e-mail messages, please include the following information, most of which can be obtained by entering `$SCHRODINGER/machid` at a command prompt:

- All relevant user input and machine output
- Glide, LigPrep, or QikProp purchaser (company, research institution, or individual)
- Primary Glide, LigPrep, or QikProp user
- Computer platform type
- Operating system with version number
- Glide version number
- LigPrep version number
- QikProp version number
- Maestro version number
- mmshare version number

Copyright Notices

C and C++ Libraries for Parsing PDB Records

The C and C++ libraries for parsing PDB records are Copyright (C) 1989 The Regents of the University of California. All rights reserved.

Redistribution and use in source and binary forms are permitted provided that the above copyright notice and this paragraph are duplicated in all such forms and that any documentation, advertising materials, and other materials related to such distribution and use acknowledge that the software was developed by the University of California, San Francisco. The name of the University may not be used to endorse or promote products derived from this software without specific prior written permission. THIS SOFTWARE IS PROVIDED “AS IS” AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE IMPLIED WARRANTIES OF MERCHANTIBILITY AND FITNESS FOR A PARTICULAR PURPOSE.

120 West 45th Street
32nd Floor
New York, NY 10036

101 SW Main Street
Suite 1300
Portland, OR 97204

3655 Nobel Drive
Suite 430
San Diego, CA 92122

Dynamostraße 13
68165 Mannheim
Germany

QuatroHouse, Frimley Road
Camberley GU16 7ER
United Kingdom

SCHRÖDINGER.